Measurements design and phenomena discrimination

Laura Rebollo-Neira
Mathematics
Aston University
Birmingham, B4 7ET, UK

February 10, 2009

Abstract

The construction of measurements suitable for discriminating signal components produced by phenomena of different nature is considered. The required measurements should be capable of cancelling out those signal components which are to be ignored when focussing on a phenomenon of interest. Under the hypothesis that the subspaces hosting the signal components produced by each phenomenon are complementary, their discrimination is accomplished by measurements giving rise to the appropriate oblique projector operator. The subspace onto which the operator should project is selected by non-linear techniques in line with adaptive pursuit strategies.

PACS numbers: 02.30.Mv, 02.60.Gf, 02.30.Sa, 83.85.Ns, 95.75.Tv, 95.75.Pq, 95.75.Fg

1 Introduction

The word *signal* is frequently used to refer to a physical carrier convening information about some phenomenon. We adopt such terminology and further refer to the process of transforming a signal into a number (within the corresponding units) as a *measurement*. An appropriate mathematical setting for this description is to consider that a signal is an element of some vector space and a measurement a functional transforming the vector into a scalar. In this effort we discuss the design of measurements in relation to the following problem: Assume that a signal f, represented as an element of an inner product space $\mathcal{H}$, arises by the superposition of two components, f_1 and f_2, each of which is produced by a particular phenomenon and such that $f_1 \in \mathcal{S}_1$ and $f_2 \in \mathcal{S}_2$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are disjoint subspaces of $\mathcal{H}$, i.e. $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$. This condition implies that the superposition $f = f_1 + f_2$ is unique. The matter to be addressed here concerns the construction of the appropriate measurements allowing us to discriminate the component, say f_1, from the available signal f and the knowledge of $\mathcal{S}_1$ and $\mathcal{S}_2$. Under the condition $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$ the problem has a straightforward 'theoretical' solution, since the component f_1 can be extracted from f by an oblique projection onto $\mathcal{S}_1$ and along $\mathcal{S}_2$ [1, 2]. Unfortunately, even when theoretically the condition $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$ is satisfied, if the subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ are not well separated, the construction of the corresponding oblique projector becomes ill posed. Consequently, the signal splitting can not be achieved by numerical calculations in finite precision arithmetics. This is the situation we are concerned with. *We assume that the given subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ are 'theoretically' disjoint, but close enough to yield an ill posed problem.*

Our proposal for the numerical realization of the phenomenon discrimination is focussed on the search of a subspace of the given $\mathcal{S}_1$, where a class of signals is considered to lie. It will be assumed throughout the paper that the class of signals to be considered is K-sparse in a spanning set for $\mathcal{S}_1$. By this we mean that given a spanning set for $\mathcal{S}_1$, the corresponding linear superposition of a signal has at most K nonzero coefficients. The K-value should be less than or equal to the dimension of the subspace of $\mathcal{S}_1$ such that the construction of measurement vectors giving rise to an oblique projection onto itself, and along $\mathcal{S}_2$, is well conditioned. This assumption is quite realistic, considering that in practice there is often a lack of complete knowledge on the actual subspace $\mathcal{S}_1$ and to be on the safe side one may overestimate it. However, the assumption does not make the problem much easier to solve. Indeed, the problem of subspace selection is in general a combinatorial problem, whereby an exhaustive search of possibilities is in general intractable. The approach we propose in this Communication evolves by step wise optimal selection and is in line with the adaptive greedy approximation termed Matching Pursuit. Such a technique, which appeared first in the statistic literature [3], has been extended in the area of signal processing to several greedy strategies [4–9] being currently of assistance to a range of disciplines, including physics [10–12]. In particular, we revise and extend the Oblique Matching Pursuit (OBMP) approach which has been recently proposed in relation the above described problem of signal discrimination [13].

The paper is organized as follows: In Section 2 we introduce the mathematical setting for signal representation to be adopted here and in Section 3 with discuss the construction of oblique projectors. Section 4 highlights the importance of the search for sparse representations in the construction of oblique projectors for phenomenon discrimination. The proposed strategy is discussed in Section 5 and illustrated in Section 6 by two numerical simulations: 1)The cancellation of impulsive noise from the register of a system of harmonic oscillators and 2) the separation of a spectrum from blackbody radiation background. The conclusions are presented in Section 7.

2 Mathematical framework

Regardless of the informational content of a signal, we deal with it mathematically by considering it as an element of an inner product space $\mathcal{H}$. Thus, adopting Dirac's notation, we represent a signal f as a ket $|f\rangle$ and the corresponding dual as a bar $\langle f|$. Accordingly, the square norm $|||f\rangle||^2$ is induced by the inner product that we indicate as $\langle f|f\rangle$. For our present purpose we further assume that all the signals of interest belong to some finite dimensional subspace $\mathcal{V} \in \mathcal{H}$ spanned by a finite set $\{|v_i\rangle \in \mathcal{H}\}_{i=1}^M$. Consequently, for every signal $|f_\mathcal{V}\rangle \in \mathcal{V}$ there exists a set of numbers $\{c_i\}_{i=1}^M$ which allow us to express the signal as the linear superposition

$$|f_\mathcal{V}\rangle = \sum_{i=1}^{M} c_i |v_i\rangle. \tag{1}$$

In the jargon of signal processing the above expansion is called *atomic decomposition* and the vectors in the decomposition are called 'atoms' [14]. In applications where an economical signal representation is important, the goal is to construct decompositions involving as few terms as possible. For this end the atoms are selected from a large and, in general redundant, set called a *dictionary*. If the number of M-terms in (1) is satisfactorily small in relation with the particular application, the decomposition is said to be *sparse*.

Even when we think of a signal as an abstract object in an inner product space, for processing tasks we need a numerical representation of such an object. The process of transforming the signal into a number is refereed to as a *measurement* or *sampling*. Since we restrict considerations to linear measurements, we represent them by *linear functionals*. Thus, making use of Riesz' theorem [15] we can express a linear measurement as $m = \langle w|f\rangle$ for some $|w\rangle \in \mathcal{H}$. Hence, considering M measurements m_i, $i = 1, \ldots, M$, each of which is obtained by a *measurement vector* $|w_i\rangle$ we have a numerical representation of the ket $|f\rangle$ as given by

$$m_i = \langle w_i|f\rangle, \quad i = 1, \ldots, M. \tag{2}$$

The representation of measures as in (2), which have been used in physics for many years, has started to become popular within other disciplines through the theory of Compressed Sensing [16–20].

The problem of reconstructing the whole information content of a signal from a numerical representation has been extensively studied for the last thirty years from a number of different points of view. The diversity of available approaches is specially helpful when the reconstruction is to be achieved on the basis of incomplete information. In particular, the above mentioned theory of Compressed Sensing has produced strong theoretical results with regard to the recovery of a signal, assumed to be sparse in some orthonormal basis, from a number of non adaptive measurements which can be significantly less than the dimension of the signal subspace [16–20]. Here we focus on the particular problem involving the reconstruction of a signal from a set of linear measurements as given in (2), with no further numerical calculations other than using these numbers in the expansion (1). In other words, *we wish to use the measurements as coefficients in the linear combination (1)*. The question then arises as to which are the conditions to be requested for the measurement vectors $|w_i\rangle$ to produce the corresponding numbers allowing the reconstruction of the signal $|f\rangle$ as

$$|f_{\mathcal{V}}\rangle = \sum_{i=1}^{M} |v_i\rangle\langle w_i|f\rangle. \tag{3}$$

A major consequence of working under the assumption that the signal of interest, $|f_{\mathcal{V}}\rangle$, belongs to a *finite* dimensional subspace, $\mathcal{V}$, is *the lack of uniqueness of the measurement vectors* $\{|w_i\rangle\}_{i=1}^M$*, even when the spanning set* $\{|v_i\rangle\}_{i=1}^M$ *is linearly independent*. This statement appears clearly from the following observation.

Let us denote $\sum_{i=1}^M |v_i\rangle\langle w_i|$ as an operator, $\hat{E}$, so as to recast (3) in the fashion

$$|f_{\mathcal{V}}\rangle = \hat{E}|f\rangle. \tag{4}$$

This equation tells us that the measurement vectors $\{|w_i\rangle\}_{i=1}^M$ should be such that operator $\hat{E}$ is a projector onto $\mathcal{V}$. Indeed, the operator $\hat{E}$ is a projector if and only if it is *idempotent* [15,21], i.e., $\hat{E}^2 = \hat{E}$. Consequently, as discussed below, the projection is onto the range of the operator, $\mathcal{R}(\hat{E})$, and along its null space $\mathcal{N}(\hat{E})$.

Denoting by $\mathcal{D}$ the domain of $\hat{E}$ we recall that

$$\mathcal{R}(\hat{E}) = \{|f\rangle, \text{ such that } |f\rangle = \hat{E}|g\rangle, \text{ for some} |g\rangle \in \mathcal{D}\}.$$

Thus, for $\hat{E}$ an idempotent operator and for $|f\rangle \in \mathcal{R}(\hat{E})$, we have $\hat{E}|f\rangle = \hat{E}^2|g\rangle = \hat{E}|g\rangle = |f\rangle$. This implies that $\hat{E}$ behaves like the identity operator for all $|f\rangle \in \mathcal{R}(\hat{E})$, regardless of $\mathcal{N}(E)$, which is defined as

$$\mathcal{N}(E) = \{|g\rangle, \text{ such that } \hat{E}|g\rangle = 0,\ |g\rangle \in \mathcal{D}\}.$$

It is now clear that to reconstruct a signal $|f\rangle \in \mathcal{V}$ by means of (3) the measurement vectors $\{|w_i\rangle\}_{i=1}^M$ should give rise to an operator $\sum_{i=1}^M |v_i\rangle\langle w_i|$, which must be a projector onto $\mathcal{V}$. It is appropriate to point out that the required operator is *not unique* (even if the spanning set $\{|v_i\rangle\}_{i=1}^M$ is linearly independent) because there exist many projectors onto $\mathcal{V}$ having different $\mathcal{N}(\hat{E})$. Consequently, for reconstructing signals in the range of the projector its null space can be chosen arbitrarily. Nevertheless, the null space, and therefore the particular measurement vectors, become crucial when the projector is to be applied on signals outside its range. It follows then that the measurement vectors $\{|w_i\rangle\}_{i=1}^M$ can be tailored for a particular purpose. Such a degree of freedom will be indicated hereforth by using two subscripts for representing a projector. We adopt the notation $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ to indicate a projector onto the subspace $\mathcal{V}$ and along the subspace $\mathcal{W}^\perp$. The particular case $\hat{E}_{\mathcal{V}\mathcal{V}^\perp}$, where $\mathcal{V}^\perp$ is orthogonal to $\mathcal{V}$, corresponds to an *orthogonal projector* and we use the special notation $\hat{P}_\mathcal{V}$ to denote such a projector. The orthogonal projector is popular in approximation techniques because if a signal $|f\rangle$ is to be approximated by a signal $|f_\mathcal{V}\rangle \in \mathcal{V}$ the choice $|f_\mathcal{V}\rangle = \hat{P}_\mathcal{V}|f\rangle$ is known to yield the unique signal in $\mathcal{V}$ minimizing the distance $|||f\rangle - |f_\mathcal{V}\rangle||$. However, as discussed below, if one is interested in discriminating signal components produced by phenomena of different nature an alternative selection of the subspace $\mathcal{W}^\perp$ is required. When $\mathcal{W}^\perp$ is not orthogonal to $\mathcal{V}$ the projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ is referred to as an *oblique projector*.

Let as assume for instance that a signal $|f\rangle$ is the superposition of two signals, $|f\rangle = |f_\mathcal{V}\rangle + |f_{\mathcal{W}^\perp}\rangle$, each component being produced by a different phenomenon we wish to discriminate. Let us assume further that we can model the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ hosting each signal component and such subspaces are disjoint, i.e. $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$. Thus we can obtain $|f_\mathcal{V}\rangle$ from $|f\rangle$, by an oblique projector onto $\mathcal{V}$ and along $\mathcal{W}^\perp$. The projector will map to zero the component $|f_{\mathcal{W}^\perp}\rangle$ to produce

$$|f_\mathcal{V}\rangle = \hat{E}_{\mathcal{V}\mathcal{W}^\perp}|f\rangle.$$

In the next section we discuss the construction of measurement vectors $\{|w_i\rangle\}_{i=1}^M$ giving rise to the desired projector.

3 Constructing measurement vectors for discrimination of signal components

Given two disjoint subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$, in order to provide a prescription for constructing the projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ one can proceed as follows. Firstly we define $\mathcal{S}$ as the direct sum of $\mathcal{V}$ and $\mathcal{W}^\perp$, which we express as

$$\mathcal{S} = \mathcal{V} \oplus \mathcal{W}^\perp.$$

Let $\mathcal{W} = (\mathcal{W}^\perp)^\perp$ be the orthogonal complement of $\mathcal{W}^\perp$ in $\mathcal{S}$. Thus we have $\mathcal{S} = \mathcal{V} \oplus \mathcal{W}^\perp = \mathcal{W} \oplus^\perp \mathcal{W}^\perp$, where the operation $\oplus^\perp$ indicates the orthogonal sum, which refers to the direct sum of orthogonal subspaces.

Considering that $\{|v_i\rangle\}_{i=1}^M$ is a spanning set for $\mathcal{V}$, a spanning set for $\mathcal{W}$ is obtained as

$$|u_i\rangle = |v_i\rangle - \hat{P}_{\mathcal{W}^\perp}|v_i\rangle = \hat{P}_\mathcal{W}|v_i\rangle,\ i = 1, \ldots, M. \quad (5)$$

Denoting as $\{|i\rangle\}_{i=1}^M$ the standard orthonormal basis for $\mathbb{C}^M$, we define the operators $\hat{V} : \mathbb{C}^M \to \mathcal{V}$ and $\hat{U} : \mathbb{C}^M \to \mathcal{W}$ as

$$\hat{V} = \sum_{i=1}^{M} |v_i\rangle\langle i|, \qquad\qquad \hat{U} = \sum_{i=1}^{M} |u_i\rangle\langle i|.$$

Thus the adjoint operators $\hat{U}^*$ and $\hat{V}^*$ are expressed as

$$\hat{V}^* = \sum_{i=1}^{M} |i\rangle\langle v_i|, \qquad\qquad \hat{U}^* = \sum_{i=1}^{M} |i\rangle\langle u_i|.$$

Notice that $\hat{P}_{\mathcal{W}}\hat{V} = \hat{U}$ and $\hat{U}^*\hat{P}_{\mathcal{W}} = \hat{U}^*$ hence, $\hat{G} : \mathbb{C}^M \to \mathbb{C}^M$ defined as:

$$\hat{G} = \hat{U}^*\hat{V} = \hat{U}^*\hat{U}$$

is a self-adjoint operator. Its matrix representation being given by the elements $\langle i|\hat{G}|j\rangle = \langle u_i|v_j\rangle = \langle u_i|u_j\rangle$, $i, j = 1, \ldots, M$.

Remark 1. *It is appropriate to stress that*

- *Operators $\hat{V}$ and $\hat{U}$ are given in terms of spanning sets for the spaces $\mathcal{V}$ and $\mathcal{W}$, respectively, and* any *such spanning set can be used.*

- *The condition $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ implies that the dimension of $\mathcal{V}$ is equal to the dimension of $\mathcal{W}$. Hence, provided that the spanning set $\{|v_i\rangle\}_{i=1}^M$ is linearly independent, operator $\hat{G}$ has an inverse. Nevertheless, the independence of $\{|v_i\rangle\}_{i=1}^M$ is not a requirement and therefore an inverse for $\hat{G}$ need not exist. For the sake of generality we shall use $\hat{G}^\dagger$, which indicates a pseudo-inverse of $\hat{G}$.*

The oblique projector operator onto $\mathcal{V}$ and along $\mathcal{W}^\perp$ is given as [22]

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{V}\hat{G}^\dagger\hat{U}^*, \tag{6}$$

or, equivalently, as

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \sum_{i=1}^{M} |v_i\rangle\langle w_i|, \tag{7}$$

with

$$|w_i\rangle = \hat{U}\hat{G}^\dagger|i\rangle = \sum_{j=1}^{M} = |u_j\rangle\langle j|\hat{G}^\dagger|i\rangle. \tag{8}$$

It is actually straightforward to verify that $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ given in (7) satisfies the required properties. Namely, $\hat{E}^2_{\mathcal{V}\mathcal{W}^\perp} = \hat{E}_{\mathcal{V}\mathcal{W}^\perp}$, $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}|f_\mathcal{V}\rangle = |f_\mathcal{V}\rangle$ for all $|f_\mathcal{V}\rangle \in \mathcal{V}$, and $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}|g\rangle = 0$ or all $|g\rangle \in \mathcal{W}^\perp$.

Remark 2. *The construction of an oblique projector is similar to that of an orthogonal one. The difference being that in general the subspaces span$\{|v_i\rangle\}_{i=1}^M = \mathcal{V}$ and span$\{|w_i\rangle\}_{i=1}^M = \mathcal{W}$ are different. For the special case $\{|v_i\rangle\}_{i=1}^M = \{|u_i\rangle\}_{i=1}^M$ we have span$\{|w_i\rangle\}_{i=1}^M =$ span$\{|v_i\rangle\}_{i=1}^M = \mathcal{V}$, thereby the projector is self adjoint and, consequently, an orthogonal projector onto $\mathcal{V}$ along $\mathcal{V}^\perp$.*

We are already in a position to extract the component $|f_\mathcal{V}\rangle$ from $|f\rangle$ by the simple operation $|f_\mathcal{V}\rangle = \hat{V}\hat{G}^\dagger\hat{U}^*|f\rangle$. However, the correct discrimination of the signal components is successful provided that the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ are well separated. Unfortunately, this is not always the case and the construction of the necessary projector may generate an ill posed problem. In spite of the fact that 'theoretically' $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, numerical errors, due to the existence of small eigenvalues values of the operator $\hat{G}$, may cause the failure to find the unique signal splitting that theoretically one should expect. Nevertheless, the correct separation is still possible, provided that the signal $|f_\mathcal{V}\rangle$ admits a sparse representation in some spanning set for $\mathcal{V}$. In order words, one could succeed in extracting $|f_\mathcal{V}\rangle$, provided that it is well represented in a subspace $\mathcal{V}_K \subset \mathcal{V}$ inducing a subspace $\mathcal{W}_K \subset \mathcal{W}$ (satisfying $\mathcal{V}_K + \mathcal{W}^\perp = \mathcal{W}_K \oplus^\perp \mathcal{W}^\perp$) where the computation of the measurement vectors is well posed. If this is the case, the problem of designing measurement vectors for discriminating signal components can be addressed as the problem of finding the subspace $\mathcal{V}_K$ where $|f_\mathcal{V}\rangle$ is well represented. Unfortunately the search for the subspace $\mathcal{V}_K$ is in general intractable. Indeed, let us assume that $\{|v_i\rangle\}_{i=1}^M$ is an spanning set for $\mathcal{V}$ and $\mathcal{V}_K$ is spanned by K elements of such a set. Even possessing this knowledge, the problem of finding the right subspace by exhaustive search would be a combinatorial problem: Out of a set of cardinality M there exist $\binom{M}{K}$ possible subsets of cardinality K. As already mentioned, we shall not look for the sparsest representation but make the search of the appropriate subspace tractable by means of recursive greedy pursuit strategies, which are only step wise optimal. Before discussing our approach some considerations are in order.

4 Getting ready for the search

In this section we highlight some properties that will be of assistance in the next section, where we will present our strategy for the search of the sparse representation achieving the desired signal discrimination. The goal is to avoid the computation of the measurement vectors in the whole subspace. Instead, we strive to find the subspace $\mathcal{V}_K \subset \mathcal{V}$, where the signal component one wants to extract from a signal $|f\rangle$ is assumed to lie. We work under the hypothesis that the subspace $\mathcal{W}^\perp$ is given and fixed. Furthermore, $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, which implies that there exists a unique solution for the signal splitting. The problem we need to address arises from the fact that, if the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ are not well separated, the numerical calculation of the measurement vectors is not accurate (due to the numerical operations being carried out in finite precision arithmetic). As a consequence, the representation of the corresponding projector fails to produce the correct signals separation. This effect is very much magnified if the data are affected by errors no matter how insignificant those errors are.

Assuming that we are able to accurately compute in finite precision arithmetic r measurement vectors, we could attempt to single out a signal belonging to a subspace spanned by at most r vectors (i.e. we could attempt to separate from $|f\rangle$ a signal expressible as in (3) but at most with r nonzero coefficients). However, as discussed above, even possessing this knowledge about the sought signal the problem of finding the right subspace by exhaustive search is not affordable. Hence, an adaptive greedy strategy for the subspace selection, given a signal, was advanced in [13]. Before revising and extending that strategy we need to recall two relevant properties of oblique projectors.

Property 1. *The oblique projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ satisfies $\hat{P}_\mathcal{W}\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \hat{P}_\mathcal{W}$.*

Proof. It readily follows by applying $\hat{P}_\mathcal{W}$ on both sides of (6) or (7). Since $\hat{P}_\mathcal{W}|v_i\rangle = |u_i\rangle$ and

$\langle u_i, v_j\rangle = \langle u_i, u_j\rangle$, one has

$$\hat{P}_{\mathcal{W}}\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \sum_{i=1}^{M} |u_i\rangle\langle w_i| = \hat{U}\hat{G}^\dagger\hat{U}^* = \hat{P}_{\mathcal{W}}. \tag{9}$$

□

Property 2. *Given a signal $|f\rangle$ in $\mathcal{V} + \mathcal{W}^\perp = \mathcal{W} \oplus \mathcal{W}^\perp$, the only vector $|g\rangle \in \mathcal{V}$ satisfying*

$$\hat{P}_{\mathcal{W}}|f\rangle = \hat{P}_{\mathcal{W}}|g\rangle \tag{10}$$

is $|g\rangle = \hat{E}_{\mathcal{V}\mathcal{W}^\perp}|f\rangle$.

Proof. If $|g\rangle = \hat{E}_{\mathcal{V}\mathcal{W}^\perp}|f\rangle$ (10) trivially follows from Property 1. Let us assume now that there exists $|g\rangle \in \mathcal{V}$ such that (10) holds. Then $\hat{P}_{\mathcal{W}}(|f\rangle - |g\rangle) = 0$, i.e., $(|f\rangle - |g\rangle) \in \mathcal{W}^\perp$. Hence $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}(|f\rangle - |g\rangle) = 0$ and, since $|g\rangle \in \mathcal{V}$, this implies that $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}|f\rangle = |g\rangle$. □

Let us suppose that $\mathcal{V}_k = \text{span}\{|v_i\rangle\}_{i=1}^k$ is given and the spanning set is linearly independent. Assuming that $\mathcal{V}_k \cap \mathcal{W}^\perp = \{0\}$ we guarantee that the set of vectors $\{|u_i\rangle\}_{i=1}^k$, with $|u_i\rangle$ given in (5), is also linearly independent. Consequently the dimension of $\mathcal{V}_k$ is equal to the dimension of $\mathcal{W}_k = \text{span}\{|u_i\rangle\}_{i=1}^k = \text{span}\{|w_i^k\rangle\}_{i=1}^k$. We use now a superscript k to indicate that the measurement vectors $\{|w_i^k\rangle\}_{i=1}^k$ span $\mathcal{W}_k$. Hence these vectors give rise to the oblique projection of a signal $|f\rangle$, onto $\mathcal{V}_k$ and along $\mathcal{W}^\perp$, as given by:

$$\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}|f\rangle = \sum_{i=1}^{k} |v_i\rangle\langle w_i^k|f\rangle = \sum_{i=1}^{k} c_i^k|v_i\rangle. \tag{11}$$

It is clear from (11) that if the atoms in the atomic decomposition were to be changed (or some atoms were added to or deleted from the decomposition) the measurement vectors $\{|w_i^k\rangle\}_{i=1}^k$, and consequently the coefficients $\{c_i^k\}_{i=1}^k$ in (11), would need to be modified. The recursive equations below provide an effective way of implementing the task.

Forward/backward adapting of measurement vectors

Starting with $|w_1^1\rangle = \frac{|u_1\rangle}{|||u_1\rangle||^2}$, and $|u_1\rangle$ as in (5), the measurement vectors $\{|w_i^{k+1}\rangle\}_{i=1}^{k+1}$ can be recursively constructed from $\{|w_i^k\rangle\}_{i=1}^k$ as follows [2]:

$$|w_i^{k+1}\rangle = |w_i^k\rangle - |w_{k+1}^{k+1}\rangle\langle u_{k+1}|w_i^k\rangle, \quad i = 1, \ldots, k \tag{12}$$

$$|w_{k+1}^{k+1}\rangle = \frac{|\gamma_{k+1}\rangle}{|||\gamma_{k+1}\rangle||^2}, \quad |\gamma_{k+1}\rangle = |u_{k+1}\rangle - \hat{P}_{\mathcal{W}_k}|u_{k+1}\rangle, \tag{13}$$

where $\hat{P}_{\mathcal{W}_k}$ is the orthogonal projector onto $\mathcal{W}_k = \text{span}\{|u_i\rangle\}_{i=1}^k$. We note that, since $|u_{k+1}\rangle = \hat{P}_{\mathcal{W}}|v_{k+1}\rangle$ and $\hat{P}_{\mathcal{W}}|w_i^k\rangle = |w_i^k\rangle$, (12) can also be written as

$$|w_i^{k+1}\rangle = |w_i^k\rangle - |w_{k+1}^{k+1}\rangle\langle v_{k+1}|w_i^k\rangle, \quad i = 1, \ldots, k. \tag{14}$$

It follows from the above equations that when incorporating a linearly independent atom $|v_{k+1}\rangle$ in the atomic decomposition (11), the coefficients can be conveniently modified according to the recursive equations

$$c_{k+1}^{k+1} = \langle w_{k+1}^{k+1}|f\rangle, \tag{15}$$

$$c_i^{k+1} = \langle w_i^{k+1}|f\rangle = c_i^k - c_{k+1}^{k+1}\langle w_i^k|v_{k+1}\rangle, \quad i = 1, \ldots, k. \tag{16}$$

Conversely, considering that the atom, $|v_j\rangle$ say, is to be removed from the atomic decomposition (11), and denoting the corresponding subspaces $\mathcal{V}_{k\setminus j}$ and $\mathcal{W}_{k\setminus j}$, in order to span $\mathcal{W}_{k\setminus j}$ the measurement vectors $\{|w_i^{k\setminus j}\rangle\}_{i=1,i\neq j}^{k}$ are modified according to the equation [2]

$$|w_i^{k\setminus j}\rangle = |w_i^k\rangle - \frac{|w_j^k\rangle\langle w_j^k|w_i^k\rangle}{|||w_j^k\rangle||^2}, \quad i = 1, \ldots, j-1, j+1, \ldots, k. \tag{17}$$

Consequently, the coefficients in (11) should be changed to

$$c_i^{k\setminus j} = c_i^k - \frac{\langle w_i^k|w_j^k\rangle c_j^k}{|||w_j^k\rangle||^2}, \quad i = 1, \ldots, j-1, j+1, \ldots, k. \tag{18}$$

5 Adaptive pursuit strategy for subspace selection

Given a signal $|f\rangle$, we aim at finding the subspace $\mathcal{V}_K \subset \mathcal{V}$ where one of the signal components lies. Let us stress once again that the problem arises from the impossibility of correctly computing the measurement vectors spanning the whole subspace $\mathcal{W}$. Moreover, we have to face the fact that the corresponding signal component we want to represent *is not available.* What we know is that the available signal, $|f\rangle$, is expressible as the sum of two components $|f\rangle = |f_{\mathcal{V}}\rangle + |f_{\mathcal{W}^\perp}\rangle$ and that there exists an unknown subspace $\mathcal{V}_K = \text{span}\{|v_{\ell_i}\rangle\}_{i=1}^K \subset \mathcal{V}$ where $\{\ell_i\}_{i=1}^K$ is a set of K unknown indexes such that $|f\rangle = |f_{\mathcal{V}}\rangle + |f_{\mathcal{W}^\perp}\rangle = |f_{\mathcal{V}_K}\rangle + |f_{\mathcal{W}^\perp}\rangle$, with

$$|f_{\mathcal{V}}\rangle = |f_{\mathcal{V}_K}\rangle = \sum_{i=1}^{K} |v_{\ell_i}\rangle\langle w_i^K|f\rangle. \tag{19}$$

Hence, if the set of indexes $\{\ell_i\}_{i=1}^K$ were given, one could construct the measurement vectors $|w_i^K\rangle$ in $\mathcal{W}_K = \text{span}\{\hat{P}_{\mathcal{W}}|v_{\ell_i}\}_{i=1}^K$ and obtain the component $|f_{\mathcal{V}}\rangle = |f_{\mathcal{V}_K}\rangle$ from (19). Unfortunately, in the problem we are addressing neither the set of indexes $\{\ell_i\}_{i=1}^K$ nor the component $|f_{\mathcal{V}}\rangle$ are given. Nevertheless, by applying $\hat{P}_{\mathcal{W}}$ to both sides of (19) we obtain

$$|f_{\mathcal{W}}\rangle = |f_{\mathcal{W}_K}\rangle = \sum_{i=1}^{K} |u_{\ell_i}\rangle\langle w_i^K|f\rangle. \tag{20}$$

Denoting $\hat{I}_{\mathcal{S}}$ to the identity operator in $\mathcal{S}$ we have $\hat{P}_{\mathcal{W}} = \hat{I}_{\mathcal{S}} - \hat{P}_{\mathcal{W}^\perp}$. Thus, since the subspaces $\mathcal{S}$ and $\mathcal{W}^\perp$ are known, we do have access to the component $|f_{\mathcal{W}}\rangle$. We can then look for the set of indexes $\{\ell_i\}_{i=1}^K$ to approximate this component as in (20).

Remark 3. *Notice that the measures $\langle w_i^K|f\rangle$ involved in* (19) *and* (20) *are the same. Therefore, by finding the representation of $|f_{\mathcal{W}_K}\rangle$ we have the information which is needed to obtain $|f_{\mathcal{V}_K}\rangle$ from* (19). *Let us stress that the need to deal with $|f_{\mathcal{W}}\rangle$ also introduces the bad conditioned nature of the problem we are considering. Certainly, having access to the signal $|f_{\mathcal{V}}\rangle$ would imply that, provided that the spanning set $\{|v_i\rangle\}_{i=1}^M$ were well conditioned, one could find the sparse approximation* (19) *without difficulty. However, even when the conditioning of this spanning set is ideal (i.e. $\{|v_i\rangle\}_{i=1}^M$ is an orthonormal basis for $\mathcal{V}$) the fact that we need to deal with the projection $|f_{\mathcal{W}}\rangle$, which is sparse in the set $\{|u_i\rangle = \hat{P}_{\mathcal{W}}|v_i\rangle\}_{i=1}^M$, introduces the difficulty we have to face. If the set $\{u_i\}_{i=1}^M$ were well conditioned, the robust signal splitting could be obtained by a simple projection. The situation we are concerned with comprises the cases in which the whole set $\{u_i\}_{i=1}^M$ is very bad conditioned but the projection* (19) *has a well conditioned representation in the subset $\{u_{\ell_i}\}_{i=1}^K \subset \{u_i\}_{i=1}^M$ we aim to find.*

The proposed strategy for selecting the subset of atoms $\{|u_{\ell_i}\rangle\}_{i=1}^K$ evolves by stepwise selection and is in line with the strategies in [23–25]. By fixing $\hat{P}_{\mathcal{W}_k}$, at iteration $k+1$ we select the index ℓ_{k+1} such that $||\hat{P}_{\mathcal{W}}|f\rangle - \hat{P}_{\mathcal{W}_{k+1}}|f\rangle||^2$ is minimized.

Proposition 1. *Let us denote by J the set of indices $\{\ell_1, \ldots, \ell_k\}$. Given $\mathcal{W}_k = span\{|u_{\ell_i}\rangle\}_{i=1}^k$, the index ℓ_{k+1} corresponding to the atom $|u_{\ell_{k+1}}\rangle$ for which $||\hat{P}_{\mathcal{W}}|f\rangle - \hat{P}_{\mathcal{W}_{k+1}}|f\rangle||^2$ is minimal is to be determined as*

$$\ell_{k+1} = \arg \max_{n \in J \setminus J_k} \frac{|\langle \gamma_n | f \rangle|}{\| |\gamma_n\rangle \|}, \; \| |\gamma_n\rangle \| \neq 0, \tag{21}$$

with $|\gamma_n\rangle$ given in (13)*, and J_k the set of indices that have been previously chosen to determine $\mathcal{W}_k$.*

Proof. It readily follows since $\hat{P}_{\mathcal{W}_{k+1}}|f\rangle = \hat{P}_{\mathcal{W}_k}|f\rangle + \frac{|\gamma_n\rangle\langle\gamma_n|f\rangle}{\|\gamma_n\|^2}$ and hence $||\hat{P}_{\mathcal{W}}|f\rangle - \hat{P}_{\mathcal{W}_{k+1}}|f\rangle||^2 = ||\hat{P}_{\mathcal{W}}|f\rangle||^2 - ||\hat{P}_{\mathcal{W}_k}|f\rangle||^2 - \frac{|\langle\gamma_n|f\rangle|^2}{\|\gamma_n\|^2}$. Because $\hat{P}_{\mathcal{W}}|f\rangle$ and $\hat{P}_{\mathcal{W}_k}|f\rangle$ are fixed, $||\hat{P}_{\mathcal{W}}|f\rangle - \hat{P}_{\mathcal{W}_{k+1}}|f\rangle||^2$ is minimized if $\frac{|\langle\gamma_n|f\rangle|}{\||\gamma_n\rangle\|}$, $\|\gamma_n\| \neq 0$ is maximal over all $n \in J \setminus J_k$. □

The OBMP selection criterion [13] selects the index ℓ_{k+1} as the maximizer over $n \in J \setminus J_k$ of

$$\frac{|\langle \gamma_n | f \rangle|}{\| |\gamma_n\rangle \|^2}, \quad ||\gamma_n|| \neq 0.$$

This condition was proposed in [13] based on the *consistency principle* [22,26]. Such a principle, introduced in [26] and extended in [22], states that the reconstruction of a signal should be self-consistent in the sense that, if the approximation is measured with the same vectors, the same measures should be obtained. Accordingly, the above OBMP criterion was derived in [13] in order to select the measurement vector $|w_{k+1}^{k+1}\rangle$ producing the maximum *consistency error* $\Delta = |\langle w_{k+1}^{k+1}|f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} f\rangle|$, with regard to a new measurement $|w_{k+1}^{k+1}\rangle$. However, since the measurement vectors are not normalized to unity, it is sensible to consider the consistency error relative to the corresponding vector norm $|||w_{k+1}^{k+1}\rangle||$, and select the index so as to maximize over $k+1 \in J \setminus J_k$ the *relative consistency error*

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}|f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} f\rangle|}{|||w_{k+1}^{k+1}\rangle||}, \quad |||w_{k+1}^{k+1}\rangle|| \neq 0. \tag{22}$$

In order to cancel this error, the new approximation is constructed accounting for the concomitant measurement vector.

Property 3. *The index ℓ_{k+1} satisfying (21) maximizes over $k+1 \in J \setminus J_k$ the relative consistency error* (22)

Proof. Since for all vector $|w_{k+1}^{k+1}\rangle$ given in (13) $\langle w_{k+1}^{k+1}|\hat{E}_{\mathcal{V}_k\mathcal{W}^\perp} = 0$ and $|||w_{k+1}^{k+1}\rangle|| = |||\gamma_{k+1}\rangle||^{-1}$ we have

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}|f\rangle|}{||w_{k+1}^{k+1}||} = \frac{|\langle \gamma_{k+1}|f\rangle|}{||\gamma_{k+1}||}.$$

Hence, maximization of $\tilde{\Delta}$ over $k+1 \in J \setminus J_k$ is equivalent to (21). □

It is clear at this point that the forward selection of indices prescribed by proposition (21) is equivalent to selecting the indices by applying the subset selection criterion introduced in [23] (c.f. Theorem 1) on the projected signal $\hat{P}_{\mathcal{W}}|f\rangle$ using the dictionary $\{|u_i\rangle\}_{i=1}^M$. In the context of subspace selection for signal representation we have termed such a criterion Optimized Orthogonal Matching Pursuit (OOMP) [6].

The hypothesis that the computation of more than r measurement vectors becomes an ill posed problem enforces the OOMP selection of indices to stop if iteration r is reached. Nevertheless, the fact that the signal is assumed to be K-sparse, with $K \leq r$, does not imply that before (or at) iteration r one will always find the correct subspace. The r-value just indicates that it is not possible to continue with the forward selection, because the computations would become inaccurate and unstable. Hence, if the right solution was not yet found, one needs to implement a strategy accounting for the fact that it is not feasible to compute more than r measurement vectors. An adequate procedure is achieved by means of the swapping-based refinement to the OOMP approach introduced in [9]. As discussed below, it consists of interchanging already selected atoms with nonselected ones.

Consider that at iteration r the correct subspace has not appeared yet and the selected indices are labeled by the r indices $\ell_1, \ldots, \ell_r$. In order to choose the index of the atom that minimizes the norm of the residual error as passing from approximation $\hat{P}_{\mathcal{W}_r}|f\rangle$ to approximation $\hat{P}_{\mathcal{W}_{r\setminus j}}|f\rangle$ we should fix the index of the atom to be deleted, ℓ_j say, as the one for which the quantity

$$\frac{|c_i^r|}{|||w_i^r\rangle||}, \; i = 1, \ldots, r. \tag{23}$$

is minimized [7, 9, 25, 27].

The process of eliminating one atom from the atomic decomposition (11) is called *backward step* while the process of adding one atom is called *forward step*. The forward selection criterion to choose the atom to replace the one eliminated in the previous step is accomplished by finding the index $\ell_i, i = 1, \ldots, r$ for which the functional

$$e_n = \frac{|\langle \nu_n | f \rangle|}{|||\nu_n\rangle||}, \quad \text{with} \quad |\nu_n\rangle = |u_n\rangle - \hat{P}_{\mathcal{W}_{r\setminus j}}|u_n\rangle, \quad |||\nu_n\rangle|| \neq 0 \tag{24}$$

is maximized. In our framework, using (17), the projector $\hat{P}_{\mathcal{W}_{r\setminus j}}$ is computed as

$$\hat{P}_{\mathcal{W}_{r\setminus j}} = \hat{P}_{\mathcal{W}_r} - \frac{\langle w_i^r | w_j^r \rangle \langle w_j^r|}{|||w_j^r\rangle||^2}.$$

Since $\hat{P}_{\mathcal{W}_r}$ and $|w_j^r\rangle$ are available, the computation of the sequence $|\nu_n\rangle$ in (24) is a simple operation.

The swapping of pairs of atoms is repeated until the swapping operation, if carried out, would not decrease the approximation error. The implementation details for an effective realization of this process are given in [9], and MATLAB codes are available at [28]. Since there is no guarantee that at the end of the swapping of pairs of atoms the correct subspace has been found, the process can continue by increasing the number of atoms the swapping involves. At the second stage, in line with [29], we propose the swapping to be realized by the combinations of two backward steps followed by two forward steps, provided that the interchange of the two atoms improves the approximation error. If at the end of the second stage the right subspace has not yet been found, the number of atoms involved in the swapping is increased up to three and so on. Notice that if the number of atoms to be interchanged reaches the value r the whole process would repeat identically. This is avoided by initiating the new circle with a different initial atom. The above specified hypothesis ensures that the algorithm will stop when the correct signal splitting has been found. At such a stage one has $\hat{P}_{\mathcal{W}}|f\rangle = \hat{P}_{\mathcal{W}_r}|f\rangle$ with $\mathcal{W}_r$ spanned by the selected atoms $\ell_1, \ldots, \ell_r$. If the order K of sparseness of the signal is less than r a number of $r - K$ coefficients in the atomic decomposition

$$|f_{\mathcal{V}_r}\rangle = \sum_{i=1}^{r} |v_{\ell_i}\rangle\langle w_i^r|f\rangle = \sum_{i=1}^{r} c_i^r |v_{\ell_i}\rangle$$

will have zero value.

In the realistic case where the measurements are affected by errors, the proposed iterative process is to be stopped when the condition

$$||\hat{P}_{\mathcal{W}}|f\rangle - \hat{P}_{\mathcal{W}_r}|f\rangle|| \leq \delta$$

is reached (where δ should be determined by taking into account the errors of the data).

6 Numerical simulations

6.1 Impulsive noise filtering

We extend here the example in [2] concerning the elimination of impulsive noise from the register of the motion of uncoupled damped harmonic oscillators.

The n-th oscillator is characterized by a frequency of $\frac{n}{2}$ Hertz and its motion, as a function of time, is given by the equation

$$\langle t|x_n\rangle = x_n(t) = e^{-t}\cos(\pi n t). \tag{25}$$

In [2] the register of system's motion was considered to be the signal

$$\langle t|f_{\mathcal{V}_{100}}\rangle = \sum_{n=1}^{100} \frac{e^{-t}\cos(\pi n t)}{(1 + 0.7(n-75)^2)}, \quad t \in [0,1] \tag{26}$$

and the impulsive noise corrupting this signal was assumed to belongs to the subspace

$$\mathcal{W}^{\perp} = \text{span}\{e^{-100000(t-0.0025j)^2}, \, j = 1, \ldots, 400, \, t \in [0,1]\}. \tag{27}$$

Due to the nature of the distribution of frequencies in (26), the corresponding oblique projector for filtering the impulsive noise from the register was recursively constructed by incrementing the frequency n one by one, until the projection of the noisy signal onto the span of the set $\{x_n(t)\}_{n=1}^{k}$ remained unaltered by increasing the number k of elements in the set up to some value. As remarked in [2], such a procedure is not always feasible. Let us consider for instance that the 100 oscillators have frequencies which are not restricted to the range $[1, 100]$ but can be any integer number in the $[1, 405]$ interval. In this case the recursive construction of the projector by incrementing the frequency one by one is in general not possible, because the corresponding numerical calculations become ill posed before reaching the frequency $n = 405$. However, the oblique projector, along $\mathcal{W}^{\perp}$ given above, onto a subspace spanned by only 100 functions $x_n(t)$, $t \in [0,1]$ with $n \in [1, 405]$, can be accurately calculated. Thus, by applying the techniques of the previous sections to find the right frequencies of the system, the cancellation of the impulsive noise in the register is rendered possible. This is illustrated by the numerical simulation described below.

The impulsive noise is simulated by randomly taking 200 pulses in (27). The system's motion is simulated as a linear combination of 100 functions (25) the frequencies of which are taking, randomly, from the set $[1, 405]$. The data are assumed to be known in single precision. The simulation was run 50 times and in all the cases the cancellation of the impulsive noise was successful. The left graph of Fig. 1 plots one of the realizations of the experiment (motion of the system plus impulsive noise vs time). The graph on the right depicts the result obtained by applying the proposed technique to the signal on the left (it coincides with the line representing the true signal). On the contrary, although the spaces (27) and $\text{span}\{x_n(t),\, t \in [0,1]\}_{n=1}^{405}$ are 'theoretically' complementary, since the construction of the corresponding oblique projector is very ill posed, the projection fails to correctly separate the signals.

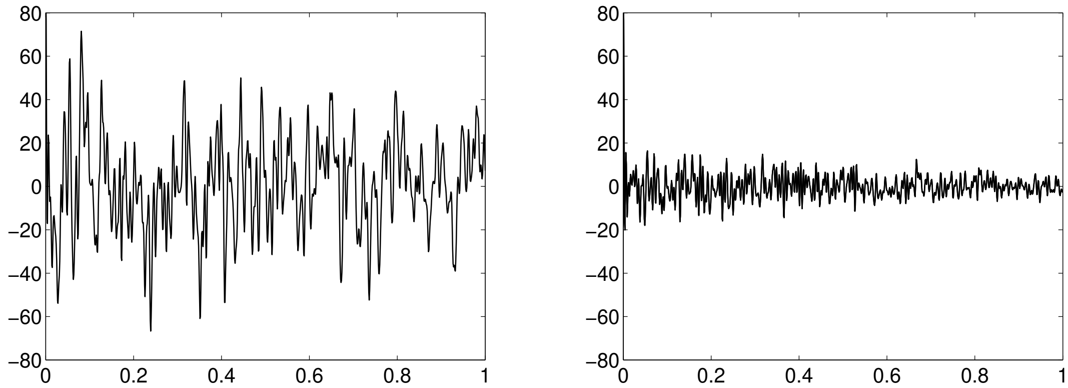

Figure 1: The left graph depicts the register of the motion of a system consisting of 100 damped harmonic oscillators, whose frequencies are integer numbers randomly taken from the interval $[1, 405]$, corrupted by 200 pulses randomly taken from the subspace $\mathcal{W}^{\perp}$ given in (27). The graph on the right depicts the result of filtering the signal of the previous graph by the strategy described in Section 5. The result coincides with the true motion of the simulated system.

6.2 Application to the separation of a spectrum from blackbody radiation background

We consider an hypothetical situation where the background is assumed to be produced by a linear combination of up to five blackbodies (e.g. stars) at temperatures of $T_1 = 3000$K, $T_2 = 3500$K , $T_3 = 4000$K, $T_4 = 4500$K , and $T_5 = 5000$K. Hence, in this case the subspace $\mathcal{W}^{\perp}$ is defined as $\mathcal{W}^{\perp} = \text{span}\{y_i\}_{i=1}^{5}$, where y_i are functions of the wavelength λ as given by

$$y_i(\lambda) = \frac{C_1}{\lambda^5(e^{\frac{C_2}{\lambda T_i}} - 1)}, \quad C_1 = 3.7419 \times 10^{-6}\text{erg}\,\text{cm}^2\,\text{s}^{-1}, C_2 = 1.4288\text{cm}\,\text{K}.$$

Any linear combination of these functions is an acceptable background. In our numerical experiment the background, $y(\lambda)$, is generated as $y(\lambda) = \sum_{i=1}^{5} y_i(\lambda)$.

We simulate a spectrum, on a region of λ ranging from zero to 3μm, by considering that it belongs to the cardinal cubic spline space on the interval $[0, 3\mu\text{m}]$ with separation $b = 2^{-4}\mu$m between consecutive knots. Such a space can be spanned by a B-spline basis arising by translating a prototype B-spline [30,31]. Different spectra are simulated by randomly drawing $K = 70$

functions $\{v_{\ell_i}\}_{i=1}^{70}$, from the basis consisting of 483 functions, to generate the decomposition $\sum_{i=1}^{70} c_i v_{\ell_i}$, with random coefficients $c_i \in [0,1]$, $i = 1\ldots,70$. A particular realization is depicted in the bottom graph of Fig. 2. The simulated available signal is obtained by adding the background $y(\lambda)$ given above to the spectrum, and perturbing each data point with a normal distributed error of variance corresponding to a percentage of each data value. Let us remark that, although the background is not 'exactly' in the cardinal spline space of the spectrum, it has a very good representation in such a space. Hence, the calculation of an oblique projector onto the given spline space and along the space of the background is expected to be very badly conditioned. Indeed, for very small errors (variance of $10^{-6}\%$ of each data value) the separation of the spectrum from the background is not possible by an oblique projection onto the whole spline space. However, in a simulation of 100 different spectra, each of which consisting of 70 randomly taken B-spline functions, the separation was successful in all the cases by applying the proposed approach. Thus, a more realistic situation was simulated by increasing the variance of the error up to 1% of each data value. Also in this case the spectrum recovery was a complete success. In order to make evident the errors' effect, the variance was increased up to 5% of each data value. One of the simulations is plotted in the top graph of Fig. 3. The middle graph of the same figure shows both, the theoretical 'hidden' spectrum and the one recovered by the proposed approach. This graph is meant to illustrate a 'typical result', as in a run of the 100 simulations described above the reconstruction of the corresponding spectra was of similar quality. The visualization of the approximation quality is made clearer in the bottom graph of Fig. 3, where a portion of the previous graph (corresponding to the interval $[0.5, 1]$) is plotted. Here we can see that, as expected, the approximation (broken line) fails to reproduce the peaks of low intensity (the one at 0.65μm). This is of course understandable by comparing the intensity of the spectrum with the intensity of the data on the $[0.5, 1]$ interval. In this region, a variance of 5% of intensity of each data point entails an uncertainty of more than one unit in the corresponding scale of intensity. Thus, one cannot expect to correctly spot peaks of intensity of the same order as the errors. On the other hand, some spurious peaks which are not in the true spectrum may also appear (note the small peak of negative intensity). However, on the whole we can confidently assert that the recovery of the simulated spectra is satisfactory even for significant error level. It is pertinent to point out that if one wished to avoid negative intensities one could penalize the selection of measurements leading to such negative values. In our framework this is implementable in a straightforward manner. For instance, in the forward selection procedure, at iteration $k+1$ we select the index ℓ_{k+1} satisfying (21) and it follows from (11) and (15) that the spectral intensity vector at this step is given as

$$|f_{\mathcal{V}_{k+1}}\rangle = |f_{\mathcal{V}_k}\rangle - \hat{E}_{\mathcal{V}_k}|v_{\ell_{k+1}}\rangle \frac{\langle \gamma_{k+1}|f\rangle}{|||\gamma_{k+1}\rangle||^2} + |v_{\ell_{k+1}}\rangle \frac{\langle \gamma_{k+1}|f\rangle}{|||\gamma_{k+1}\rangle||^2}, \tag{28}$$

where $|f_{\mathcal{V}_k}\rangle$ is the spectral intensity obtained in the previous iteration, $\hat{E}_{\mathcal{V}_k}$ the oblique projector onto the previously selected subspace, $|f\rangle$ the observed data vector and $|\gamma_{k+1}\rangle$ (constructed as in (13)) is to be determined in the selection process of the index ℓ_{k+1} according to the prescription of Proposition 1. Thus, restrictions on $|f_{\mathcal{V}_{k+1}}\rangle$ can be incorporated by disregarding the selected indices yielding unacceptable values of $|f_{\mathcal{V}_{k+1}}\rangle$. In the example we are discussing here, one can avoid negative values of intensity by disregarding those indices which satisfying (21) do not fulfill the condition $\langle\lambda|f_{\mathcal{V}_{k+1}}\rangle = f_{\mathcal{V}_{k+1}}(\lambda) \geq 0$ for the values of λ being considered. With the incorporation of this constrain the small negative pick in the middle graph of Fig. 3 disappears and the whole approximation improves. However, in some other realizations of the experiment when introducing the possibility constrains some small spurious picks of positive intensity appear, yielding on the whole an approximation of quality comparable with the unconstrained one. The spurious small peaks (or the absence or small peaks present in the true spectrum) are consequence of the considerable uncertainty in the data. The fidelity of the approximation with the true spectrum (in all the realizations of the experiment) is improved by reducing the error of the data.

7 Conclusion

The construction of measurement vectors specially designed for separating signal components produced by phenomena of different nature was discussed. Assuming that the subspaces hosting the signal components are given, the required measurement vectors should yield an oblique projection along one of the subspaces and onto the other. Considerations were restricted to those cases for which such subspaces are theoretically complementary, yet very close to each other, so that the construction of the measurement vectors for the whole space renders an ill posed problem. A recursive strategy for finding the right subspace to achieve the desired signal separation was then discussed. By recourse to numerical simulations it was illustrated that, provided that the signal is sparse in a spanning set of the signal subspace, the required signal splitting may be achieved by means of adaptive greedy techniques capable of searching for the required subspace while maintaining stability in the calculations. When tested in situations involving significant level of errors the proposed technique produced satisfactory results. Therefore, we are led to conclude that the framework for measurement design advanced here should be of assistance to a variety of applications where the discrimination of phenomena of different nature is required.

Acknowledgements

Support from EPSRC (EP/D062632/1) is acknowledged.

Figures

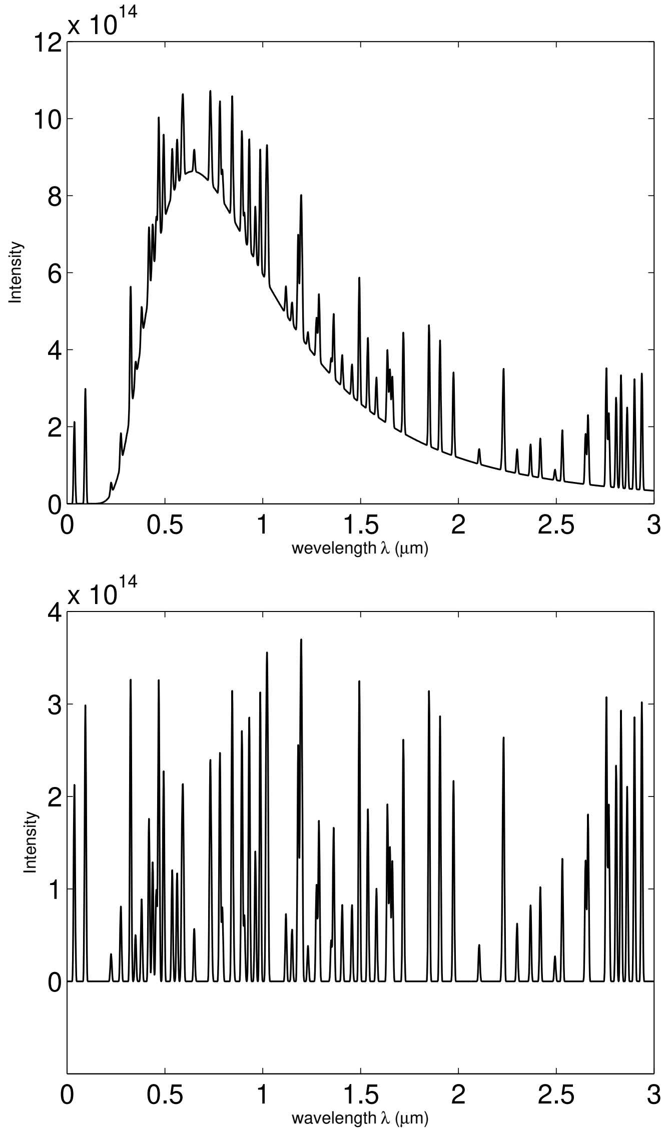

Figure 2: The top graph depicts theoretical data produced by the superposition of the spectrum shown in the bottom graph and black body radiation background.

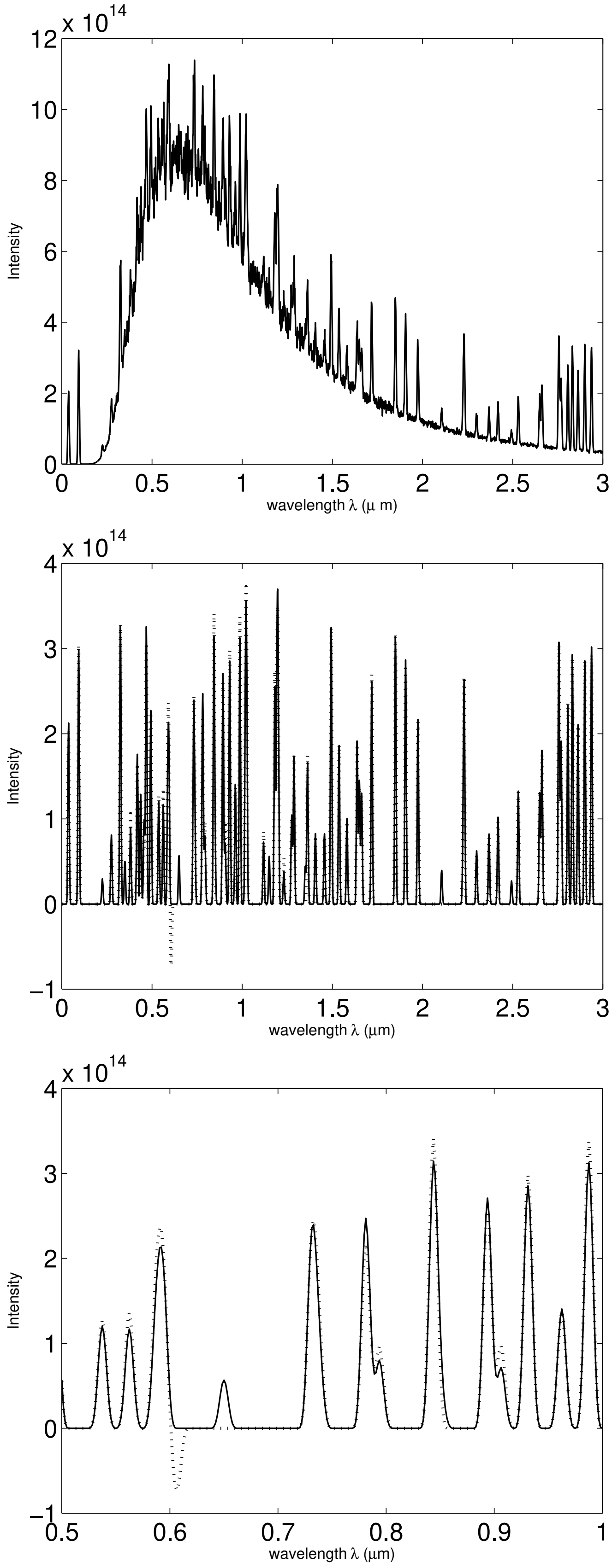
x 10^14
Intensity
wavelength λ (μ m)
x 10^14
Intensity
wavelength λ (μm)
x 10^14
Intensity
wavelength λ (μm)

Figure Caption

Figure 3. The top graph has the same description as the top graph of Fig. 2, but the data are perturbed with zero mean normal distributed errors of variance corresponding to 5% of each data value. The continuous line in the middle graph represents the theoretical spectrum and the dotted line the spectrum obtained by the proposed approach from the data of the top graph. The bottom graph magnifies the region [0.5 1] in the previous graph.

References

[1] R. Behrens, L. Scharf, Signal processing applications of oblique projection operators, IEEE Transactions on Signal Processing 42 (1994) 1413–1424.

[2] L. Rebollo-Neira, Constructive updating/downdating of oblique projectors: a generalization of the Gram–Schmidt process, Journal of Physics A: Mathematical and Theoritcal 40 (2007) 6381–6394.

[3] K. Jones, On a conjecture of huber concerning the convergence of projection pursuit regression, Ann. Stat. 15 (2) (1987) 880–882.

[4] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (1993) 3397–3415.

[5] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Constructive Approximation 13 (1) (1997) 57–98.

[6] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Processing Letters 9 (2002) 137–140.

[7] M. Andrle, L. Rebollo-Neira, E. Sagianos, Backward-optimized orthogonal matching pursuit approach, IEEE Signal Proc. Let. 11 (2004) 705–708.

[8] J. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Transactions on Information Theory 50 (10) (2004) 2231–2242.

[9] M. Andrle, L. Rebollo-Neira, A swapping-based refinement of orthogonal matching pursuit strategies, Signal Processing 86 (2006) 480–495.

[10] Y. Wu, V. S. Batista, Matching-pursuit for simulations of quantum processes, Journal of Chemical Physics 118 (2003) –.

[11] Y. Wu, V. S. Batista, Quantum tunneling dynamics in multidimensional systems, Journal of Chemical Physics 121 (2004) –.

[12] Q. Jie, Approximating the ground state of fermion systems by multiple determinant states: Matching pursuit approach, Phys. Rev. E 77 (2008) 026705.

[13] L. Rebollo-Neira, Oblique matching pursuit, IEEE Signal Processing Letters 14 (10) (2007) 703–706.

[14] S. Mallat, A Wavelet Tour of Signal Processing, 2nd Edition, Academic Press, London, 1999.

[15] M. Reed, B. Simon, Functional Analysis, Academic Press, London, 1980.

[16] D. Donoho, Compressed sensing, IEEE Trans. on Information Theory 52 (2006) 1289–1306.

[17] E. Candès, J. Romberg, Quantitative robust uncertainty principles and optimally sparse decompositions, Foundations of Comput. Math. 6 (2006) 227–254.

[18] E. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. on Information Theory 52 (2006) 5406–5425.

[19] R. Baraniuk, A lecture on compressive sensing, IEEE Signal Processing Magazine.

[20] Compressive sensing resources, http://www.dsp.ece.rice.edu/cs/See references listed there.

[21] A. Galantai, Projectors and Projection Methods, Kluwer Academic Publishers, London, 2004.

[22] Y. Eldar, Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors, Journal of Fourier Analysis and Applications 9 (2003) 77–96.

[23] L. Rebollo-Neira, A. Plastino, Nonextensive maximum-entropy-based formalism for data subset selection, Phys. Rev. E 65 (1) (2002) –.

[24] L. Rebollo-Neira, A. Plastino, A recursive approach for constructing the $q = 1/2$ maxent distribution from redundant data, Phys. Rev. E 66 (3) (2002) –.

[25] L. Rebollo-Neira, A. Plastino, Constructive approximations of the $q = 1/2$ maximum entropy distribution from redundant and noisy data, Phys. Rev. E 70 (1) (2004) –.

[26] M. Unser, A. Aldroubi, A general sampling theory for nonideal acquisition devices, IEEE Trans. Signal Processing 42 (1994) 2915–2925.

[27] L. Rebollo-Neira, Backward adaptive biorthogonalization, International Journal of Mathematics and Mathematical Science 2004 (35) (2004) 1843–1853.

[28] M. Andrle, L. Rebollo-Neira, Biorthogonal techniques for optimal signal representation, http://www.ncrg.aston.ac.uk/Projects/BiOrthog.

[29] M. Andrle, L. Rebollo-Neira, Improvement of orthogonal matching pursuit strategies by backward and forward movements, in: Proc. of the 31st International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 2006, pp. III–III.

[30] L. Schumaker, Spline Functions: Basic Theory, Wiley, New York, 1981.

[31] M. Andrle, L. Rebollo-Neira, Cardinal B-spline dictionaries on a compact interval, Applied and Computational Harmonic Analysis 18 (2005) 336–346.